**Rajesh G**

**Azure Bigdata Engineer**

**Email**: [rg.azure7@gmail.com](mailto:rg.azure7@gmail.com)

**Profile**:

* Having 10 years of experience in Development, Testing, Analysis and Implementation of Data warehousing Technology by using **Azure Data Factory, Azure Data Lake, Azure SQL, DataBricks , Azure Synapse Analytics ,Pyspark, Teradata, Informatica,Python.**
* Experience in building data pipelines using **Azure data factory** which extract the data from on premises databases ,files to transform the data using **DataBricks** and load the data into Azure SQL .
* Experience in developing application using Hadoop ecosystem like Spark, HDFS, Yarn, Hive, Sqoop.
* Having good knowledge on **AWS and Snowflake**
* Worked on Teradata MLoad, Fload, TPT utilites to load the data from files into Teradata database.
* Good understanding of Spark Architecture, MPP Architecture, including Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks.
* Using Sqoop, exported the data from Hadoop to SQL Server database.
* Worked on different file formats like Avro, parquet files in Hadoop.
* Having experience on Teradata database, worked on BTEQ scripts and Macros to process the data
* Widely used the Informatica Power Center Client Tools like Designer, Workflow Manager, Workflow Monitor and Repository Manager.
* Implemented SCD type1 and type2 using Informatica power center.
* Involved in Informatica and Teradata performance tuning.
* Well versed exposure on RDBMS like Oracle and SQL Server and Teradata.
* Worked in Agile process, sprint planning using Microsoft TFS and jira.
* Strong Interpersonal Skills, Self-motivation, Co-ordination Skills.

**Technical Expertise**

|  |  |
| --- | --- |
|  |  |

Big Data Tools Hadoop, HDFS, Hive, Sqoop

Cloud Services Azure Data Factory, Synapse Analytics, Databricks, Snowflake

Data Warehouse/ETL Tool Informatica PowerCenter, Streamsets

Databases Teradata, Oracle 12c, Sql server

Languages SQL, Pyspark, Python

**Education**:

Master of Computer Applications (MCA) from Osmania University (India).

**Professional Experience**:

**Company: R2 technologies USA**

**Client: Anthem.Inc, GA July2022- Till date**

**Role: Software developer.**

Anthem, Inc. is working to transform health care with trusted and caring solutions. Its health plan companies deliver quality products and services that give their members access to the care they need.

This Project builds the Data Warehouse which provides services to analyze the CMS stars ratings.

Responsibilities:

* Extract, Transform and Load data from Sources Systems to Azure Data Storage services and Azure SQL using Azure Data Factory. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL) and processing the data using Databricks.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, write-back tool and backwards.
* Worked on migration of data from On-prem Teradata to Cloud databases (Azure SQL DB).
* Build Complex distributed systems involving huge amount data handling, collecting metrics building data pipeline, and Analytics.
* Worked on Azure Synapse Analyticsand Databases.
* Extracted, transformed and loaded the data into Azure Synapse Analytics.
* Have good experience in working with Azure BLOB and Data Lake storage.

**Environment:** Azure Data Factory, Azure Data Lake Storage, Azure SQL, Azure Synapse Analytics, Hadoop, Hive, Spark, Python.

**Company: Legato health technologies**

**Client: Anthem.Inc, GA Dec 2020 -Dec 2021**

**Role: Sr Software developer.**

Anthem, Inc. is working to transform health care with trusted and caring solutions. Its health plan companies deliver quality products and services that give their members access to the care they need.

This Project builds the Data Warehouse which provides services to analyze the health survey data with member data.

Responsibilities:

* Extract, Transform and Load data from Sources Systems to Azure Data Storage services and Azure SQL using Azure Data Factory. Data Ingestion to one or more Azure Services - (Azure Data Lake, Azure Storage, Azure SQL) and processing the data using Databricks.
* Created Pipelines in ADF using Linked Services/Datasets/Pipeline/ to Extract, Transform and load data from different sources like Azure SQL, Blob storage, write-back tool and backwards.
* Worked on migration of data from On-prem Teradata to Cloud databases (Azure SQL DB).
* Build Complex distributed systems involving huge amount data handling, collecting metrics building data pipeline, and Analytics.
* Worked on Azure Synapse Analyticsand Databases.
* Extracted, transformed and loaded the data into Azure Synapse Analytics.
* Have good experience in working with Azure BLOB and Data Lake storage.
* Worked on Pyspark to process the data from Hive tables .
* Worked on TDCH to export and import the data from Hadoop to Teradata .

**Environment:** Azure Data Factory, Azure Data Lake Storage, Hadoop, Hive, Spark, Python, Yarn,Teradata.

**Company: Artech Infosystems India**

**Client: Dell Technologies Texas Sep 2019 – Aug 2020**

**Role: software developer**

Dell is an American [multinational](https://en.wikipedia.org/wiki/Multinational_corporation) computer technology company that develops, sells, repairs, and supports computers and related products and services.

This project Data Warehouse provides services to Orders department. This Real time data warehouse would lead the business users in analyzing the risk factors and profitability by effectively tracking Orders.

Responsibilities:

* As a software developer, worked on Streamsets application to process the real time data from Oracle data base and Hadoop file system.
* Worked on Teradata macros to process the data.
* Worked on Teradata TPT utility to load the data from file into Teradata table..
* Worked on informatica power center to extract the data from source , process the data and load the data into Teradata table.
* Used mapping parameters and variables.
* Worked on different transformations like filter,expression, transaction control transformation, aggregator to process the data.
* Involved in Unit testing.

**Technology:** Streamsets, Unix , Oracle 12C,Teradata,Informatica Power center,Hadoop.

**Appulse Technologies India Nov 2017 – Aug 2019**

**ETL developer**

Colruyt Group is a Belgian family business which, over three generations, has grown into a retail group , active in three countries. And the group still has the strengths of a family business, such as passion, expertise, simplicity and efficiency. Our aim is sustainable growth, with respect for health, society, animal welfare and the environment.

Responsibilities:

* Worked on Informatica PowerCenter tools like Source Analyzer, Target Designer to import the source and target database schemas and worked on Mapping Designer, Workflow Manger and Workflow Monitor.
* Extensively worked in data Extraction, Transformation and loading from source to target system using BTEQ,TPT.
* Created jobs on control m scheduler.
* Used transformations like Aggregators, Sorter, lookups, Filters, Expression, Router, Joiner, Source Qualifier, Update Strategy, Sequence Generator
* Used Teradata Mload connection in informatica to load the data from Oracle database to Teradata database.

**Environment:** Informatica, Oracle, Unix, Teradata.

**Tetrasoft pvt ltd India Oct 2016 – Oct 2017**

**software developer**

Anthem, Inc. is working to transform health care with trusted and caring solutions. Its health plan companies deliver quality products and services that give their members access to the care they need.

This Project builds the Data Warehouse which provides services to Clinical Program. Using this Data warehouse Business users can analyze Claims in their business

Responsibilities:

* Worked on Informatica PowerCenter tools like Source Analyzer, Target Designer to import the source and target database schemas and worked on Mapping Designer, Workflow Manger and Workflow Monitor.
* Extensively worked in data Extraction, Transformation and loading from source to target system using Teradata Bteq, and TPT.
* Used transformations like Aggregators, Sorter, lookups, Filters, Expression, Router, Joiner, Source Qualifier, Update Strategy, Sequence Generator
* Used Teradata Mload connection in informatica to load the data from Oracle database to Teradata database.
* Good understanding of implementation phase.
* Identified Bottlenecks in the mapping, and Involved in Performance tuning.
* Involved in unit testing

**Environment**: Informatica Power Center 9.5, Teradata, Unix, control m

**Birlasoft pvt ltd India April 2012 – Mar 2016**

**Software Developer**

GE Capital offers consumers and businesses around the globe an array of financial products and services. For over 1 million businesses, it provides real estate, equipment and inventory financing; fleet services; working capital; and funds for corporate acquisitions, re-financings and restructurings. GE Capital offers credit cards, retail sales finance programs, home, car and personal loans and credit insurance. GE Capital is a leading provider of specialist finance throughout the world.

This Project builds the Data Warehouse which provides services to finance department. Using this Data warehouse Business users can analyze the profits from their business.

Responsibilities:

* Worked on Informatica PowerCenter tools like Source Analyzer, Target Designer to import the source and target database schemas and worked on Mapping Designer, Workflow Manger and Workflow Monitor.
* Used transformations like Aggregators, Sorter, lookups, Filters, Expression, Router, Joiner, Source Qualifier, Update Strategy, Sequence Generator
* Used Teradata Mload connection in informatica to load the data from Oracle database to Teradata database.
* Worked on Teradata Bteqs to process the etl transformations to build the data mart.

**Environment:** Informatica Power Center, Oracle, Teradata, Unix